LLNL-PROC-415049

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Adaptive Waveform Correlation Detectors for Arrays: Algorithms for Autonomous Calibration

F. Ringdal, D. B. Harris, D. Dodge, S. J. Gibbons

July 24, 2009

Monitoring Research Review
Tucson, AZ, United States
September 21, 2009 through September 23, 2009

ADAPTIVE WAVEFORM CORRELATION DETECTORS FOR ARRAYS: ALGORITHMS FOR AUTONOMOUS CALIBRATION

Frode Ringdal[1], David B. Harris[2,3], Doug Dodge[2], Steven J. Gibbons[1]

[1]NORSAR, [2]Lawrence Livermore National Laboratory, and [3]Deschutes Signal Processing LLC

Sponsored by National Nuclear Security Administration
Contract No. DE-FC52-05NA26604[1] and DE-AC52-07NA28116

Proposal number BAA05-16

ABSTRACT

Waveform correlation detectors compare a signal template with successive windows of a continuous data stream and report a detection when the correlation coefficient, or some comparable detection statistic, exceeds a specified threshold. Since correlation detectors exploit the fine structure of the full waveform, they are exquisitely sensitive when compared to power (STA/LTA) detectors. The drawback of correlation detectors is that they require complete knowledge of the signal to be detected, which limits such methods to instances of seismicity in which a very similar signal has already been observed by every station used. Such instances include earthquake swarms, aftershock sequences, repeating industrial seismicity, and many other forms of controlled explosions. The reduction in the detection threshold is even greater when the techniques are applied to arrays since stacking can be performed on the individual channel correlation traces to achieve significant array gain.

In previous years we have characterized the decrease in detection threshold afforded by correlation detection across an array or network when observations of a previous event provide an adequate template for signals from subsequent events located near the calibration event. Last year we examined two related issues: (1) the size of the source region calibration footprint afforded by a master event, and (2) the use of temporally incoherent detectors designed to detect the gross envelope structure of the signal to extend the footprint. In Case 1, results from the PETROBAR-1 marine refraction profile indicated that array correlation gain was usable at inter-source separations out to one or two wavelengths. In Case 2, we found that incoherent detectors developed from a magnitude 6 event near Svalbard were successful at detecting aftershocks where correlation detectors derived from individual aftershocks were not. Incoherent detectors might provide "seed" events for correlation detectors that then could extend detection to lower magnitudes.

This year we addressed a problem long known to limit the acceptance of correlation detectors in practice: the labor intensive development of templates. For example, existing design methods cannot keep pace with rapidly unfolding aftershock sequences. We successfully built and tested an object-oriented framework (as described in our 2005 proposal) for autonomous calibration of waveform correlation detectors for an array. The framework contains a dynamic list of detectors of several types operating on a continuous array data stream. The list has permanent detectors: beamforming power (STA/LTA) detectors which serve the purpose of detecting signals not yet characterized with a waveform template. The framework also contains an arbitrary number of subspace detectors which are launched automatically using the waveforms from validated power detections as templates. The implementation is very efficient such that the computational cost of adding subspace detectors was low. The framework contains a supervisor that oversees the validation of power detections, and periodically halts the processing to revise the portfolio of detectors. The process of revision consists of collecting the waveforms from all detections, performing cross-correlations pairwise among all waveforms, clustering the detections using correlations as a distance measure, then creating a new subspace detector from each cluster. The collection of new subspace detectors replaces the existing portfolio and processing of the data stream resumes. This elaborate scheme was implemented to prevent proliferation of closely-related subspace detectors.

The method performed very well on several simple sequences: 2005 "drumbeat" events observed locally at Mt. St. Helens, and the 2003 Orinda, CA aftershock sequence. Our principal test entailed detection of the aftershocks of the San Simeon earthquake using the NVAR array; in this case, the system automatically detected and categorized approximately 2/3 of the events above magnitude 2.8.

OBJECTIVE

Correlation detection is becoming a mainstream option for network operations due to its advantages in sensitivity and event screening. Correlation detectors wrap detection, location and event identification functions into a single operation, which potentially makes them effective at reducing the burden of analysts in network monitoring operations. Since such detectors both detect and classify, they support a strategy of reviewing repeating events from particular sources as aggregated groups rather than one at a time. This feature may lead to significant efficiencies during aftershock sequences and swarms and in regions with large amounts of mining activity.

However, the development of correlation detectors is labor intensive as currently practiced. Typically, large numbers of events must be assembled from catalogs or by running power detectors over continuous data streams. Cross-correlations among the event waveforms are computed and events are grouped by a clustering algorithm, which brings together events with significant waveform similarity. Waveforms from the clustered events then are selected to define a correlation template, which can be applied as a matched filter to the data in a continuous stream. Subspace detectors, which are higher-dimensional extensions of correlation detectors require careful alignment of waveforms from multiple events in a cluster, construction of an orthonormal basis for the event waveforms and selection of an optimum basis dimension for signal representation. In return, they provide greater scope for signal representation.

As these detector design activities currently are manual or only semi-automated, it is not possible to keep up with the occurrence of swarm events or aftershock sequences. However, one of the most attractive potential applications of correlation detectors is as a real-time screen for the very large number of similar events that can overwhelm network operations.

It also is the case that a large number of sources surround many stations, requiring distinct, dedicated detectors. The number of detectors required is compounded by the fact that the detailed structure of signals may change over time for many of these sources, requiring detector updates.

Our solution to these problems is to automate correlation detector development to the extent possible, eventually under analyst review. The objective in seeking automation is not to replace analysts, but rather to assist them by organizing detected events into categories prior to review. We anticipate that even a partial organization of events can substantially reduce the burden of event formation and review by analysts.

RESEARCH ACCOMPLISHED

Our principal accomplishment this year was the construction and testing of a detection framework for autonomous calibration of correlation and subspace detectors (Figure 1). The heart of the system is a list of detectors that operate on a continuous stream of array data. The list contains a set of fixed STA/LTA (power) detectors implemented on a collection of beams (to allow beam recipes to be implemented for an array) that serve to obtain waveforms of new event types to spawn into correlation detectors. Correlation and subspace detectors are added dynamically to the list as described below. The system acquires a block of array data from the continuous stream, preprocesses (filters and decimates) it, then directs each detector in the list to calculate a detection statistic from the block of data. The system examines the statistic for excursions above a predetermined threshold, at which point a trigger is declared. When two or more detectors produce simultaneous triggers, the triggers are compared, and only one is promoted as a detection subject to the following rules:

(1) Triggers from the same type of detector (array power, or correlation /subspace) are promoted or eliminated based on which has the largest detection statistic value.

(2) Triggers from correlators and subspace detectors always are promoted over those from power detectors. All detections are archived to a database with information about the detector that originated them, the trigger time and the value of the detection statistic.

Detections from power detectors are assumed to be signals not yet seen: the system passes such detected waveforms through a series of screens (e.g. duration, bandwidth) in an attempt to eliminate spikes and other unwanted signal types. Waveforms that pass these tests are used to create correlation detectors which are added it to the detector list. The system then continues on to the next block of data.

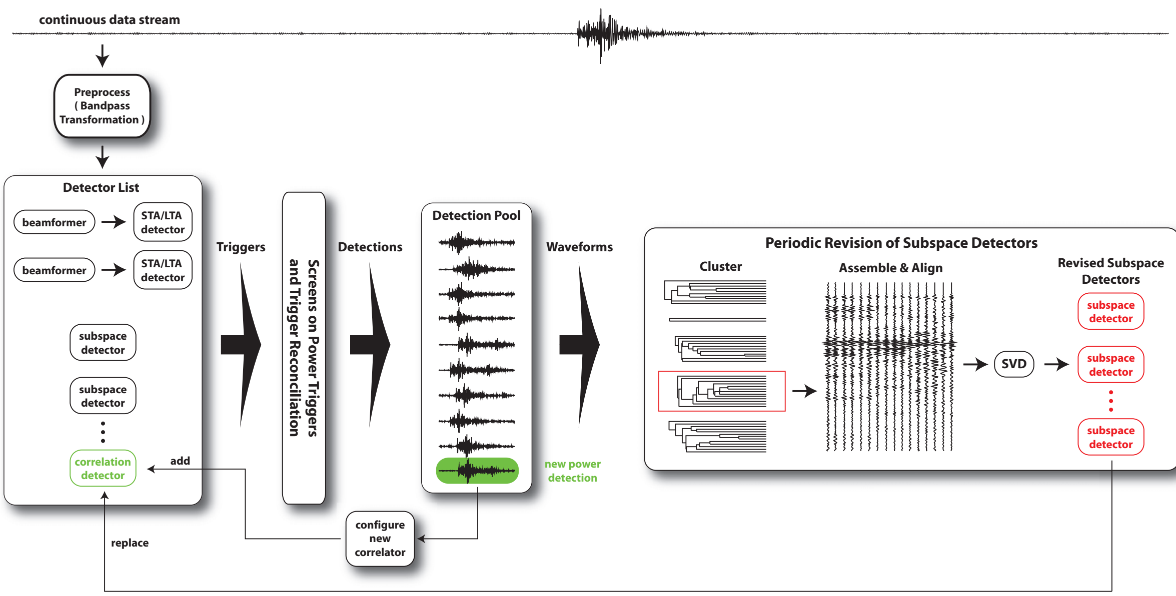

Figure 1. Diagram of the autonomous framework built to test the concept of automated construction of correlation detectors.

Periodically the system halts (after reaching a threshold in numbers of new detections) to recalibrate the correlation detectors. The purpose of this operation is to provide some check on the (otherwise unrestricted) growth in the number of correlation detectors. During recalibration, the new detections and some older detections are grouped to define new templates. Corresponding waveforms are extracted from the archive (detection pool in Figure 1). Correlations are calculated between the waveforms for all pairs of these events. The correlation values are used to cluster the events and the correlation lags are used to align event waveforms from individual clusters. A subspace detector template is constructed from the aligned waveforms from each cluster, and the collection of subspace detectors so created replaces the correlation detectors in the detector list. At this point the system continues from the point where it left off in the stream.

2003 San Simeon Earthquake

For a suitable test of this system, we sought an event that generated large numbers of aftershocks in a short period of time, endowed good ground truth information, and observed by an array at regional distance with many high-SNR observations suitable for generating correlation templates. We settled on the 2003 San Simeon earthquake (Figure 2). This was a moderately large event (mb 6.5) with thousands of aftershocks recorded by local networks in California. We acquired ten days of data (2003:356 - 2003:365) from the NVAR array for the test. NVAR is situated 390 kilometers from the main shock location. For ground truth information, we relied upon the Advanced National Seismic System (ANSS) composite catalog [NCEDC, 2009] which reported 1433 events in a 1x1 degree square around the main shock during the last ten days of 2003.

Our objective in performing this test was not to reduce the detection threshold (as is a common objective with correlation detectors), but rather automatically to group detected events as an aid to analyst review. By this measure, a system is successful if it automatically classifies a large fraction of events as they occur or with periodic bulk processing during a sequence, without introducing large numbers of unwanted detections that are difficult to review. The system we implemented was able to process the 10 days of NVAR data (9 channels @ 40 samples/second) in 15 to 20 minutes on a modern laptop computer. For ease in catalog reconciliation, and to increase the number of detections, we added a second processing step in which the 98 subspace detectors created during the first pass were held fixed and used to reprocess the entire ten days of data. This reprocessing step took an additional ten minutes. Since the system is so fast, periodic reprocessing of weeks of data is possible in real-time monitoring operations.

We assume that, with such a system, analysts could review events in groups automatically determined to be related on the basis of waveform correlation. Our results suggest that this approach is possible, and could form the basis of an improvement to current practice in network operations. We imagine a scenario where, immediately following the occurrence of a large event, an analyst directs the system to deploy a beam to the slowness of the main shock, anticipating aftershocks. The system then attempts to detect and cluster aftershocks for subsequent review. By reviewing events in groups, time and effort could be minimized, as only one or a few of the largest events in the cluster would require intensive interpretation, and the remainder could be treated as local related events of similar origin.

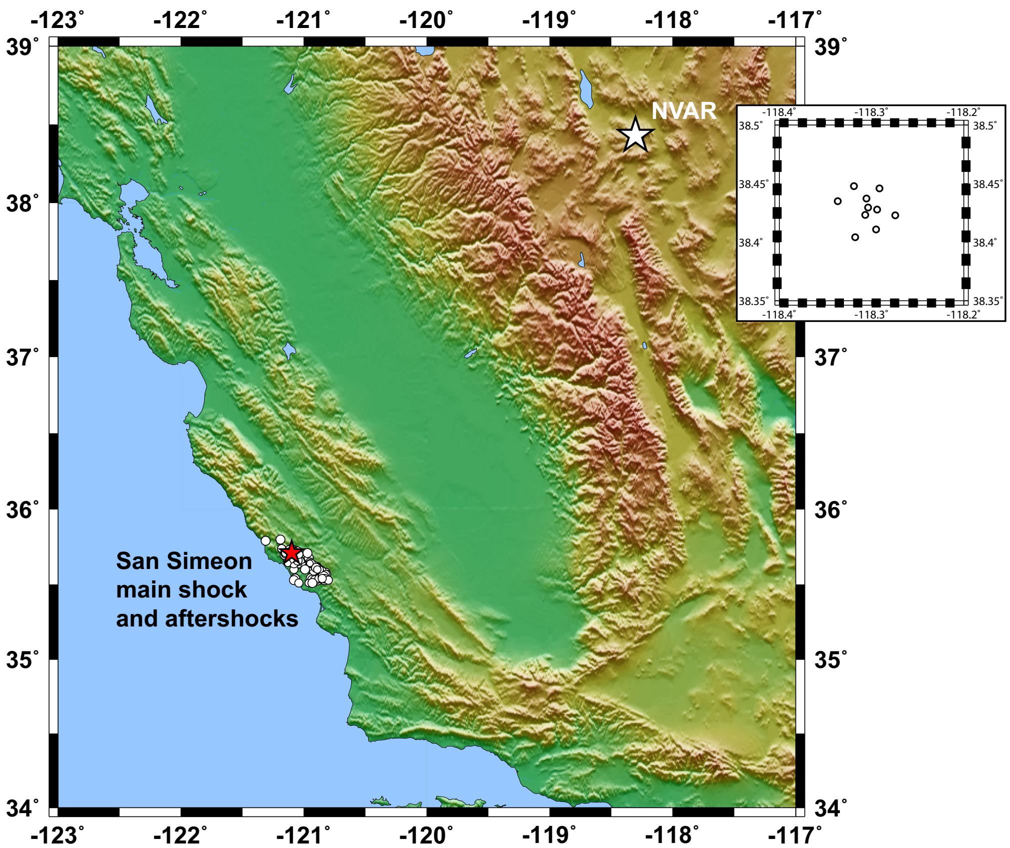

Figure 2. The main sequence of study consists of the magnitude 6.5 San Simeon earthquake and its aftershocks during the 10 day period Dec 22, 2003 - Dec 31, 2003. We acquired 10 days of data from the NVAR array to use in the test.

In our implementation, we used a standard beamformer with 9 elements of the NVAR array (2 of the 11 elements had very significant problems with dropouts). We used the great circle path backazimuth of 220 degrees and a velocity of 8 km/sec as the beamforming parameters. An STA/LTA detector was used on the beam to make power detections. The STA duration was 5 seconds, the LTA duration was 50 seconds and we inserted a 5 second gap between the STA and LTA windows. The detection threshold was set at 5 (in amplitude, 25 in power) in order to obtain high-quality signals for correlation templates.

We set the system to halt and recalibrate every time 200 new detections were made. We made distinctions between two classes of waveform correlation detectors: first generation and second generation. The first generation detectors were those created directly from STA/LTA detections (i.e. the correlators). Second generation detectors were those created from a recalibration episode and could be subspace detectors of higher rank. A database of detections was maintained with each detection “linked” to its originating detector. During a reclustering interlude all new detections of any type and all detections from first generation detectors not previously reclustered were assembled into a secondary pool. The pool events then were correlated pairwise and reclustered using a single link algorithm with a correlation clustering threshold of 0.5. Subspace detectors were developed for each identified cluster, using an energy capture metric to define the dimension of the subspace [Harris, 2006]. Events used to construct the new second-generation detectors were reassigned in the database to these new detectors. Detectors left with no detections following reassignment were removed from the system. Second generation detectors were assumed to be mature: their detections were not subject to reclustering.

This rather complex system was designed to control the number of detectors operating in the system. The number would tend to grow without bound without some process to identify and retire old detectors, replacing them with merged detectors created as new events reveal linkages among event groups. Our hypothesis was that natural groupings among the aftershocks of the San Simeon sequence and other events observed by the array would be revealed only as events accumulate over time. In our view, the system needed to be flexible to allow detectors generated early in the process to be replaced as more events allowed better correlation and subspace templates to be developed.

Over the course of processing the 10 days of data, the STA/LTA algorithm spawned 124 correlation detectors directly. The system halted and reclustered three times, ultimately reducing the number of correlation/subspace detectors to 98 (plus the beamformer). As mentioned earlier, we reprocessed the data completely a second time (starting over from the beginning) using the detectors in existence at the end of the first pass. The system was cleared of all detection records and entirely new detections were declared during the second pass. This operation had the effect of putting detections from a single detector on a common relative timing basis, which greatly simplified reconciliation against the ANSS catalog. It also increased the number of correlation and subspace detections, since templates developed from high-SNR events occurring late in the 10 day period matched smaller events occurring earlier in the sequence that were overlooked in the first pass.

Overall, the second reprocessing step produced 702 detections, of which 360 could be reconciled against the ANSS catalog. The locations of the detected events are indicated as red symbols in Figure 3 against the background of grey symbols representing the ANSS catalog. Note that most of the larger events were captured by the system and that the detected events were distributed over most of the aftershock region.

Figure 4 compares histograms of ANSS catalog events and our detections against event magnitude. These plots indicate that the system captured the majority of events above magnitude 2.8. We believe that we missed a number of the higher magnitude events due to the large duration of our long-term average in the beamforming detector. Many of these events occurred early in the sequence when large numbers of events packed together, loading the LTA and reducing the STA/LTA ratio below the detection threshold. By missing the additional templates provided by these high magnitude events, we also lost many lower magnitude events that might otherwise been detected.

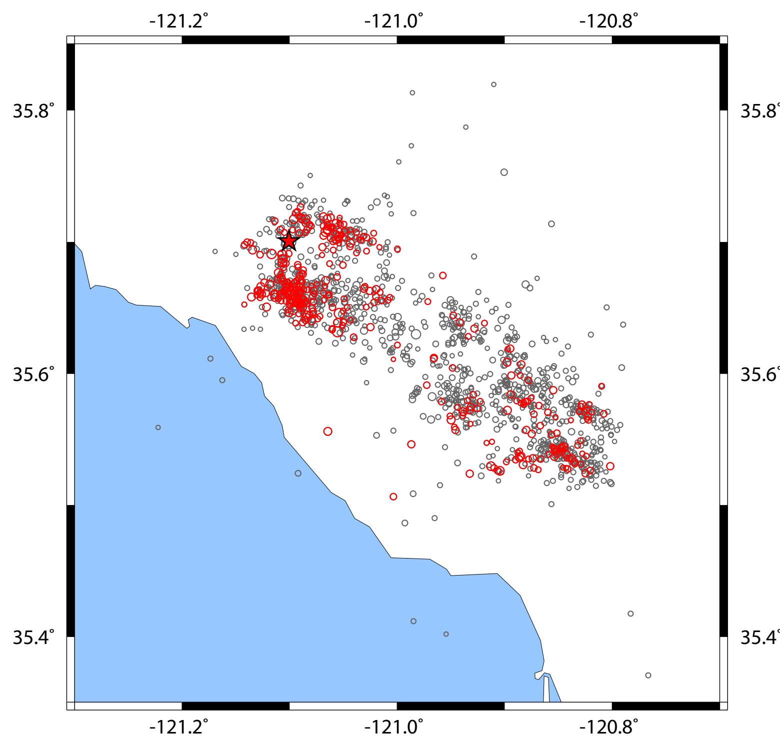

Figure 3. A map of the ANSS catalog events in the vicinity of the San Simeon main shock (grey) and the events automatically detected by our detection framework (red).

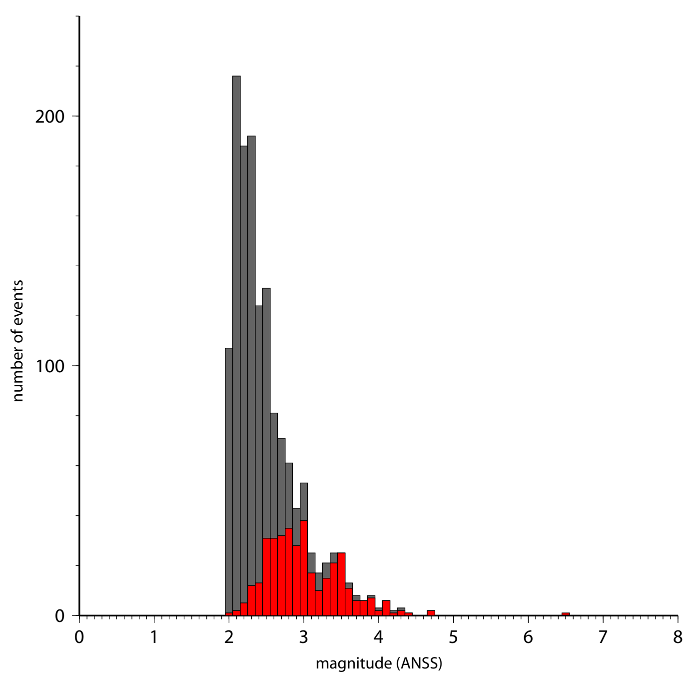

Figure 4. Histograms of the ANSS catalog events against magnitude (grey) and the events detected by our system (red). The system captured the majority of events above magnitude 2.8.

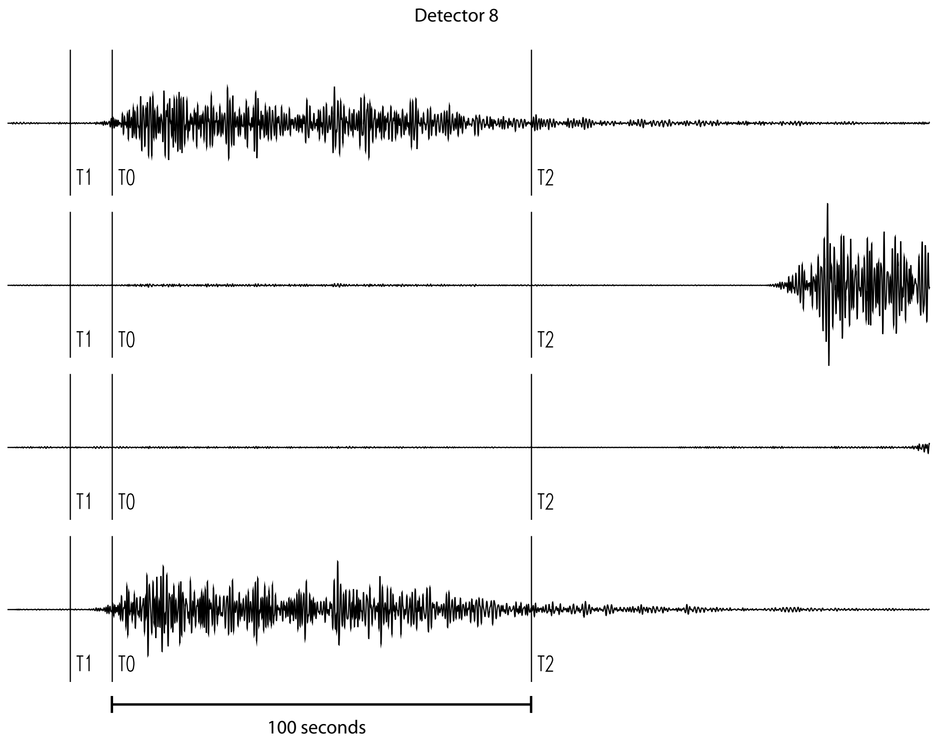

Figure 5. Example waveforms of detections associated with one detector (#8). The window used to create detectors is 110 seconds long and delimited by markers labeled T1 and T2. T0 is the nominal detection point (actually the point of detection by the STA/LTA algorithm). Common scaling of the traces causes the middle two events, which were smaller, to be invisible. The similarity of the traces allows multiple events to be interpreted simultaneously.

The key issue is whether the events are clustered in some fashion that supports a reduction in analyst effort. Figure 5 shows four detections from one detector, and demonstrates that the waveforms can be aligned automatically for comparison based on waveform correlation measurements. Several of the clusters have a large number of events (Figure 6) which should aid rapid interpretation of the sequence.

An issue of major concern is whether automatically spawned detectors introduce a large number of spurious detections, which could overwhelm any advantage gained by organizing aftershocks into compact groups for interpretation. We note that the San Simeon aftershocks occurred in 75 clusters containing 360 detections. The other 342 detections were split among 23 clusters. These detections were predominantly legitimate seismic detections from sources other than the San Simeon sequence. The largest two clusters of these detections are shown in Figure 7. These events (in fact most of the other detections) appear to be very local signals that may have entered the system through a sidelobe of the array processor. They are relatively easy to distinguish and occur in large groups that would simplify their interpretation and dismissal.

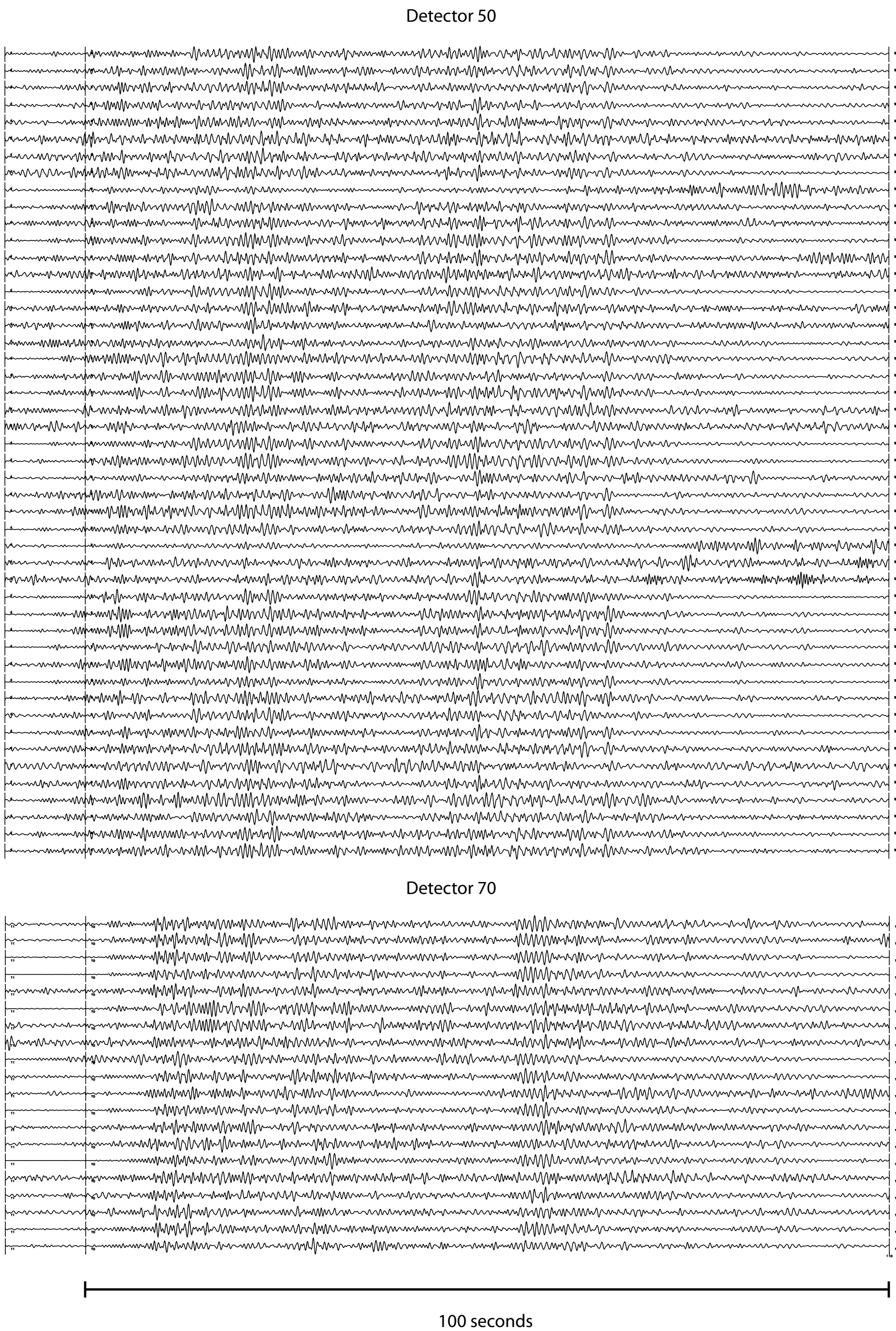

Figure 6. Two of the largest clusters of San Simeon aftershocks, automatically assembled and aligned, support our contention that efficient review by analysts would be enabled by future implementations of systems like this one.

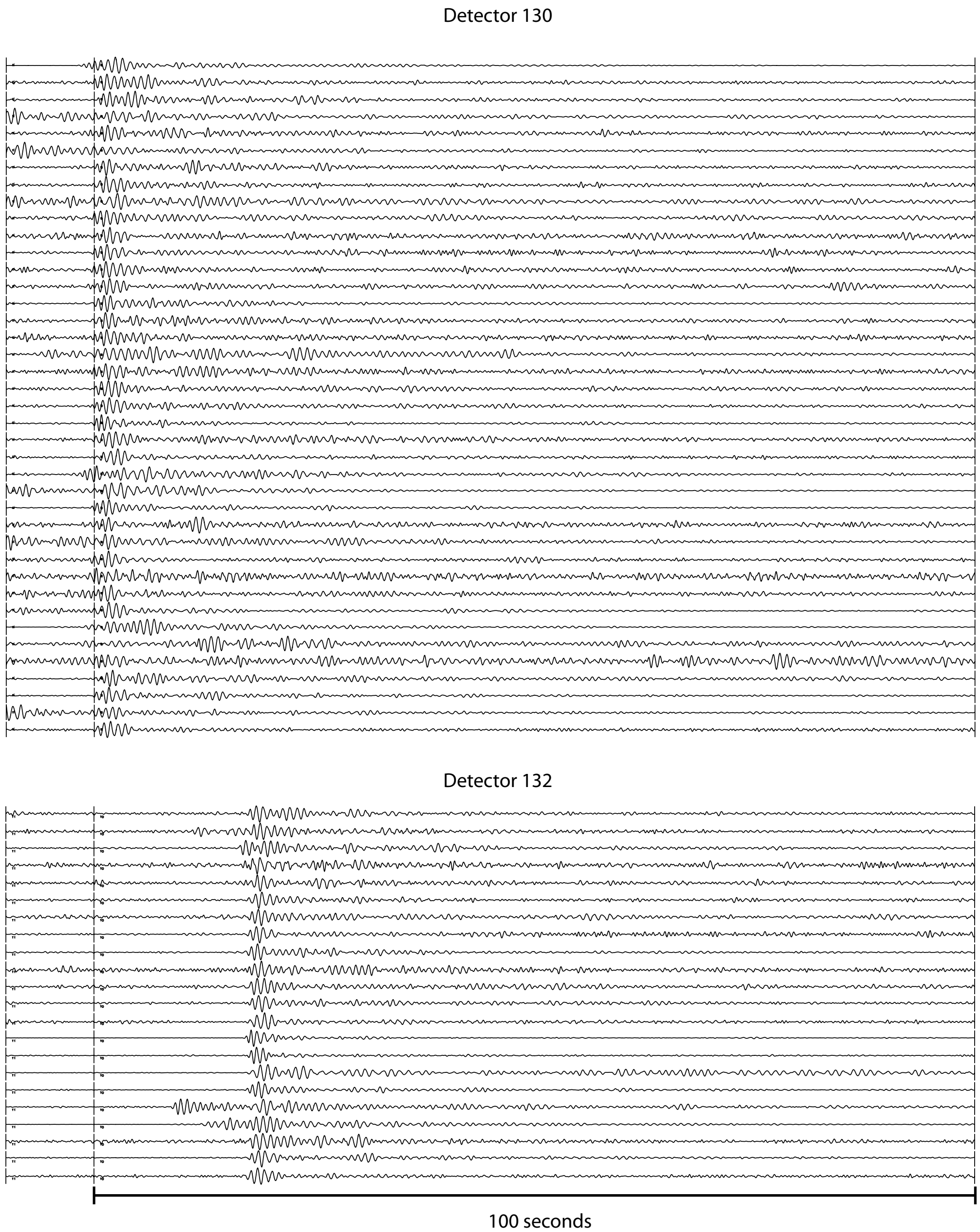

Figure 7. Two of the larger clusters of detections unrelated to the San Simeon sequence suggest that, while the system detects other local events generally not of interest, the majority of these are easily dismissed in automatically constructed groups.

CONCLUSIONS AND RECOMMENDATIONS

Overall we believe this work demonstrates that self-calibrating, autonomous correlation detection frameworks are feasible. The system we describe is a pilot; it is clear that many improvements are possible. In particular, significantly more attention should be paid to the power detector used to originate waveform correlation templates. Our system missed a significant number of the higher magnitude events, in part due to inter-event interference early in the sequence. The loss of templates from the larger events has a cascading effect, potentially eliminating large numbers of lower magnitude detections. Window selection is another area of potential improvement. We used a particularly simple algorithm, defining the template window to start and end with fixed offsets from the STA/LTA detection point. This choice of window propagates thereafter into all generations of correlation detectors. A better approach might examine windows about the detection point for temporal coherence among events within a cluster to optimize window selection. The availability of many snapshots of the same signature could be used to advantage in defining the template window. Finally, though our system did not create large numbers of spurious detections which were difficult to dismiss, potential for mischief in automatic detector creation remains. One way to suppress this problem is to extend the construction of correlation detectors to networks of stations or arrays. A network-wide correlator may have the effect of suppressing interfering signal sources local to individual stations.

REFERENCES

Harris, D. B. (2006). “Subspace Detectors: Theory”, Lawrence Livermore National Laboratory, internal report UCRL-TR-222758, 46 pp.

Harris, D. B. and Paik, T. (2006). “Subspace Detectors: Efficient Implementation”, Lawrence Livermore National Laboratory, internal report UCRL-TR-223177, 38 pp.

NCEDC (2009), ANSS catalog website, www.ncedc.org/anss/

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.